Exp Mail EV335610972US
USAN 10/816,459

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
3 July 2003 (03.07.2003)

PCT

(10) International Publication Number
WO 03/054232 A2

(51) International Patent Classification[7]: C12Q 1/68

(21) International Application Number: PCT/US02/40138

(22) International Filing Date:
12 December 2002 (12.12.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/341,409 13 December 2001 (13.12.2001) US

(71) Applicant *(for all designated States except US)*: **BLUE HERON BIOTECHNOLOGY, INC.** [US/US]; 22310, 20th Avenue Southeast, Bothell, WA 98021 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **TABONE, John, C.** [US/US]; 12117 NE 166th Place, Bothell, WA 98011 (US). **MULLIGAN, John, T.** [US/US]; 5823, 17th Avenue Northeast, Seattle, WA 98105 (US).

(74) Agents: **SHARKEY, Richard, G.** et al.; Seed Intellectual Property Law Group PLLC, Suite 6300, 701 Fifth Avenue, Seattle, WA 98104-7092 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 03/054232 A2

(54) Title: METHODS FOR REMOVAL OF DOUBLE-STRANDED OLIGONUCLEOTIDES CONTAINING SEQUENCE ERRORS USING MISMATCH RECOGNITION PROTEINS

(57) Abstract: In embodiments of the present invention, methods are provided for removing double-stranded oligonucleotide (*e.g.*, DNA) molecules containing one or more sequence errors, generated during nucleic acid synthesis, from a population of correct oligonucleotide duplexes. In one embodiment, the oligonucleotides are generated enzymatically. Heteroduplex (containing mismatched bases) oligonucleotides may be created by denaturing and reannealing the population of duplexes. The reannealed oligonucleotide duplexes are contacted with a mismatch recognition protein that interacts with (*e.g.*, binds and/or cleaves) the duplexes containing a base pair mismatch. The oligonucleotide heteroduplexes that have interacted with such a protein are separated, simultaneously with contacting or sequentially in a separate step, from homoduplexes. These methods are also used in another embodiment to remove heteroduplex oligonucleotides (*e.g.*, DNA) that are formed directly from chemical nucleic acid synthesis.

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
3 July 2003 (03.07.2003)

PCT

(10) International Publication Number
WO 03/054232 A2

(51) International Patent Classification[7]: C12Q 1/68

(21) International Application Number: PCT/US02/40138

(22) International Filing Date:
12 December 2002 (12.12.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/341,409 13 December 2001 (13.12.2001) US

(71) Applicant *(for all designated States except US)*: **BLUE HERON BIOTECHNOLOGY, INC.** [US/US]; 22310, 20th Avenue Southeast, Bothell, WA 98021 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **TABONE, John, C.** [US/US]; 12117 NE 166th Place, Bothell, WA 98011 (US). **MULLIGAN, John, T.** [US/US]; 5823, 17th Avenue Northeast, Seattle, WA 98105 (US).

(74) Agents: **SHARKEY, Richard, G.** et al.; Seed Intellectual Property Law Group PLLC, Suite 6300, 701 Fifth Avenue, Seattle, WA 98104-7092 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 03/054232 A2

(54) Title: METHODS FOR REMOVAL OF DOUBLE-STRANDED OLIGONUCLEOTIDES CONTAINING SEQUENCE ERRORS USING MISMATCH RECOGNITION PROTEINS

(57) Abstract: In embodiments of the present invention, methods are provided for removing double-stranded oligonucleotide (*e.g.*, DNA) molecules containing one or more sequence errors, generated during nucleic acid synthesis, from a population of correct oligonucleotide duplexes. In one embodiment, the oligonucleotides are generated enzymatically. Heteroduplex (containing mismatched bases) oligonucleotides may be created by denaturing and reannealing the population of duplexes. The reannealed oligonucleotide duplexes are contacted with a mismatch recognition protein that interacts with (*e.g.*, binds and/or cleaves) the duplexes containing a base pair mismatch. The oligonucleotide heteroduplexes that have interacted with such a protein are separated, simultaneously with contacting or sequentially in a separate step, from homoduplexes. These methods are also used in another embodiment to remove heteroduplex oligonucleotides (*e.g.*, DNA) that are formed directly from chemical nucleic acid synthesis.

METHODS FOR REMOVAL OF DOUBLE-STRANDED OLIGONUCLEOTIDES CONTAINING SEQUENCE ERRORS USING MISMATCH RECOGNITION PROTEINS

BACKGROUND OF THE INVENTION

Field of the Invention

The present invention in certain embodiments is directed toward the removal of double-stranded oligonucleotides containing sequence errors. It is more particularly related to the removal of error-containing oligonucleotides (such as error-containing double-stranded DNA), generated for example by chemical or enzymatic synthesis (including by PCR amplification), by removal of mismatched duplexes using mismatch recognition proteins.

Description of the Related Art

For purposes of this application, DNA is used as a prototypical example of an oligonucleotide. Mismatches are formed directly during chemical DNA synthesis or are formed in enzymatically synthesized DNA by denaturing and reannealing a mixed population of correct and error-containing DNA.

In chemical DNA synthesis, the mismatches originate during the synthesis of oligonucleotides ("oligos"). These oligos are used as building blocks for DNA synthesis and are synthesized as single strands using automated oligonucleotide synthesizers. Random chemical side reactions create base errors in these single-stranded oligos. When two complementary synthetic oligos are hybridized to form double-stranded DNA, there is almost no chance that the random base errors formed in one strand will be correctly base paired in the opposite strand. It is these incorrectly paired bases that form the mismatches found in chemically synthesized double-stranded DNA.

In enzymatic DNA synthesis, an enzyme (such as a polymerase) is used to amplify or assemble from a synthetic DNA template. This template contains the same type of base mismatches that are found in the synthetic DNA described above. However, once this DNA is amplified, the mismatches are converted into base paired errors in sequence. These base pairings of the mismatches occur as polymerase synthesizes the complementary base on the strand opposing strand. The result of this enzymatic step is to create a mixed population of DNA molecules where all bases are paired correctly with both correct (error-free) and incorrect (error-containing) sequences. The polymerase step essentially maintains the ratio of correct to incorrect sequence.

A DNA population such as that formed from enzymatic DNA synthesis containing both error-free and error-containing base paired DNA where both are correctly base pair matched, can be converted to a population composed of both mismatched and error-free correctly base paired DNA by denaturation and reannealing. When these steps are performed on a population that contains a small fraction of error-containing molecules relative to correct molecules, the vast majority of error containing strands will hybridize with the more abundant correct strand and will form mismatched sites.

Moreover, even if the errors represent a high fraction of the population (*e.g.*, 50%) denaturation and reannealing of a DNA population to itself, will result in the vast majority of a particular error-containing strand hybridizing either to a correct strand or to a strand that contains a distinct error. Thus, a population of DNA will be converted into two populations of mostly base paired correct DNA. The correct strands will find correct strand complementary strands and form perfectly base paired duplexes.

Due to the difficulties in the current approaches to the preparation or amplification of oligonucleotides, such as genes, there is a need in the art for methods for improving the removal of double-stranded oligonucleotides containing sequence errors. The present invention fills this need, and further provides other related advantages.

BRIEF SUMMARY OF THE INVENTION

Briefly stated, in certain embodiments the present invention provides a variety of methods for removing double-stranded oligonucleotide (*e.g.*, DNA) molecules containing one or more sequence errors generated during nucleic acid synthesis, from a population of correct oligonucleotide duplexes. In one embodiment, the oligonucleotides are generated enzymatically. Heteroduplex oligonucleotides may be created by denaturing and reannealing the population of duplexes. The reannealed oligonucleotide duplexes are contacted with a mismatch recognition protein that interacts with the duplexes containing a base pair mismatch. The oligonucleotide heteroduplexes that have interacted with the protein are separated from homoduplexes as the latter do not interact with the protein. These methods are also used to remove heteroduplex oligonucleotides (*e.g.*, DNA) that are formed directly from chemical nucleic acid synthesis.

In one embodiment, the present invention provides a method of depleting in a sample of double-stranded oligonucleotides a population of double-stranded oligonucleotides containing mismatched bases thereby enriching in said sample a population of double-stranded oligonucleotides containing correctly matched bases, comprising the steps of: (a) contacting said sample containing double-stranded oligonucleotides with a mismatch recognition protein under conditions to permit the

protein to interact with a double-stranded oligonucleotide containing at least one mismatched base; and (b) collecting double-stranded oligonucleotides that have not interacted with said mismatch recognition protein, thereby depleting the population of double-stranded oligonucleotides containing mismatched bases. In another embodiment, there is, prior to the step of collecting, an additional step comprising separating said double-stranded oligonucleotide containing at least one mismatched base that has interacted with said mismatch recognition protein, from double-stranded oligonucleotides that have not interacted with said mismatch recognition protein.

These and other aspects of the present invention will become evident upon reference to the following detailed description. In addition, various references are set forth herein. Each of these references is incorporated herein by reference in its entirety as if each was individually noted for incorporation.

DETAILED DESCRIPTION OF THE INVENTION

Prior to setting forth the invention, it may be helpful to an understanding thereof to set forth definitions of certain terms to be used hereinafter.

Natural bases of DNA – adenine (A), guanine (G), cytosine (C) and thymine (T). In RNA, thymine is replaced by uracil (U).

Synthetic double-stranded oligonucleotides – two strands of oligonucleotides (*e.g.*, substantially double-stranded DNA) composed of single strands of oligonucleotides synthetically produced (*e.g.*, by chemical synthesis or by the ligation of synthetic double-stranded oligonucleotides to other synthetic double-stranded oligonucleotides to form larger synthetic double-stranded oligonucleotides) and joined together in the form of a duplex.

Synthetic failures - undesired products of oligonucleotide synthesis; such as side products, truncated products or products from incorrect ligation.

Side products – chemical byproducts of oligonucleotide synthesis.

Truncated products – all possible shorter than the desired length oligonucleotide, *e.g.*, resulting from inefficient monomer coupling during synthesis of oligonucleotides.

TE – an aqueous solution of 10 mM Tris and 1 mM EDTA, at a pH of 8.0.

Homoduplex oligonucleotides – double-stranded oligonucleotides wherein the bases are fully matched; *e.g.*, for DNA, each A is paired with a T, and each C is paired with a G.

Heteroduplex oligonucleotides – double-stranded oligonucleotides wherein the bases are mispaired, *i.e.*, there are one or more mismatched bases; *e.g.*, for DNA, an A is paired with a C, G or A, or a C is paired with a C, T or A, etc.

Mismatch recognition protein – a protein that recognizes heteroduplex oligonucleotides (*e.g.*, heteroduplex DNA); typically the protein is a mismatch repair enzyme or other oligonucleotide binding protein (*e.g.*, DNA mismatch repair enzyme or other DNA binding protein); the protein may be isolated or prepared synthetically (*e.g.*, chemically or enzymatically), and may be a derivative, variant or analog, including a functionally equivalent molecule which is partially or completely devoid of amino acids.

The present invention is directed in certain embodiments toward methods for the removal of error-containing double-stranded oligonucleotide (*e.g.*, DNA) molecules from a population of double-stranded oligonucleotides (*e.g.*, that are produced by chemical or enzymatic synthesis). The error-containing oligonucleotide molecules in this population are removed from the correct molecules when the errors are present as mismatches in the double-stranded oligonucleotides. The removal of the mismatch is based in the present invention on the use of mismatch recognition proteins that recognize mismatched bases in double-stranded oligonucleotides. Such proteins interact with double-stranded oligonucleotides containing mismatched bases (*e.g.*, by binding and/or cleaving on or near the mismatch site). The protein step may or may not be performed in conjunction with a separation step (*e.g.*, chromatographic step) to separate mismatch-containing heteroduplex from homoduplex oligonucleotides. It is to be understood that the methods of the invention have the capability of mismatch removal regardless of the way the mismatch was created in the population.

More specifically, the disclosure of the present invention shows surprisingly that mismatch recognition proteins may be used to deplete an oligonucleotide population of those double-stranded oligonucleotides which contain sequence errors. Depletion of error-containing oligonucleotides from the desired double-stranded oligonucleotides refers generally to at least about (wherein "about" is within 10%) a two-fold depletion relative to the total population prior to separation. Typically, the depletion will be a change of about two-fold to three-fold from the original state. The particular fold depletion may be the result of a single use of the method (*e.g.*, single separation) or the cumulative result of a plurality of use (*e.g.*, two or more separations). Depletion of error-containing oligonucleotides is useful, for example, where the oligonucleotides are double-stranded DNA which correspond to a gene or fragments of a gene.

Oligonucleotides suitable for use in the present invention are any double-stranded sequence. Examples of such oligonucleotides include double-stranded

DNA, double-stranded RNA, DNA/RNA hybrids, and functional equivalents containing one or more non-natural bases. Preferred oligonucleotides are double-stranded DNA. Double-stranded DNA includes full length genes and fragments of full length genes. For example, the DNA fragments may be portions of a gene that when joined form a larger portion of the gene or the entire gene.

As noted above, the present invention provides a preparative method to remove base mismatched oligonucleotides from a population of correctly base matched oligonucleotides. The method generally comprises the steps of contacting a double-stranded oligonucleotide sample with a mismatch recognition protein, and collecting the double-stranded oligonucleotides that have not interacted with the mismatch recognition protein. Collecting the double-stranded oligonucleotides that have not interacted with the protein can be the result of their removal from the sample, or the removal from the sample of those oligonucleotides that did interact. The step of contacting is performed under conditions (including a time sufficient) to permit a mismatch recognition protein to interact with (*e.g.*, bind to and/or cleave) mismatch-containing heteroduplex oligonucleotides. The method may, prior to the step of collecting, optionally include a step of separating the double-stranded oligonucleotide that contains at least one (one or more) mismatched base and that has interacted with the mismatch recognition protein, from double-stranded oligonucleotides that have not interacted with the mismatch recognition protein. The method may, in place of or in addition to a separation step and prior to the step of contacting, optionally include steps of first denaturing and then reannealing a sample of double-stranded oligonucleotides under conditions to permit conversion of the double-stranded oligonucleotides first to single-stranded oligonucleotides and then to double-stranded oligonucleotides. It will be evident to one of ordinary skill in the art that the steps may be performed sequentially, or two or more steps may be performed simultaneously. For example, in an embodiment where a mismatch recognition protein is immobilized on a solid support, the step of contacting results directly in separation.

In one embodiment the mismatch recognition proteins share the property of binding on or within the vicinity of a mismatch. Such a protein reagent includes proteins that are endonucleases, restriction enzymes, ribonucleases, mismatch repair enzymes, resolvases, helicases, ligases and antibodies specific for mismatches. Variants of these proteins can be produced, for example, by site directed mutagenesis, provided that they are functionally equivalent for mismatch recognition. The enzyme can be selected, for example, from T4 endonuclease 7, T7 endonuclease 1, S1, mung bean endonuclease, MutY, MutS, MutH, MutL, cleavase, and HINF1. In another embodiment of the invention,

the mismatch recognition protein cleaves at least one strand of the mismatched DNA in the vicinity of the mismatch site.

The optional separation step can be performed in a variety of means, *e.g.*, using high performance liquid chromatography (HPLC), by size exclusion chromatography, ion exchange chromatography, affinity chromatography or reverse phase chromatography. The separation can also be performed using membranes in a slot blot fashion or a microtiter filter plate. The separation may also be performed using solid phase extraction cartridges using supports similar to the HPLC columns.

In one embodiment, a mismatch recognition protein (*e.g.*, the MutS protein from *E. coli*) is immobilized on a solid support. Methods for immobilizing proteins on solid supports are well known to one in the art, and include covalent or noncovalent attachment to a solid support. Similarly, types of suitable solid supports are well known to one in the art, and include beads, glass, polymers, resins and gels. The following is a representative example for preparing oligonucleotides depleted of error-containing oligonucleotides. Two complementary oligonucleotides (*e.g.*, DNA) are chemically synthesized and then hybridized to form duplex oligonucleotides (*e.g.*, double-stranded DNA). Alternatively, double-stranded DNA may be enzymatically synthesized (and further denatured and reannealed). This mixture is passed over a column with a mismatch recognition protein (*e.g.*, the MutS protein) immobilized on a solid support (such as beads) in the column. Fragments with an error in either of the oligonucleotides will usually contain a mismatch since in most cases the other strand is correct at that position. Duplexes containing mismatches will bind to the column and only error-free duplexes will be enriched in the flow-through from the column.

In another embodiment, a gene encoding a mismatch recognition protein (*e.g.*, the MutS gene) is fused to a gene fragment that encodes a binding domain (for instance a chitin-binding domain). The following is another representative example for preparing oligonucleotides depleted of error-containing oligonucleotides. The fused protein is produced and mixed with a duplex fragment that is produced as described above. Duplex molecules with an error in either strand will bind to the fusion protein (*e.g.*, MutS fusion protein). After an appropriate incubation, the mixture is passed over a chitin column. The fusion protein binds to the column via the chitin. Duplex molecules with mismatches are retained on the column, and error-free duplexes flow through.

The following examples are offered by way of illustration and not by way of limitation.

EXAMPLES

EXAMPLE 1

SYNTHESIS OF A 205 BP DNA FRAGMENT FROM THE OPERATOR-BINDING REGION OF THE LACI GENE

Beta-galactosidase is an enzyme that can convert X-gal from a colorless compound into a brilliant blue compound (Manniatis; Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratories, Cold Spring Harbor, NY, 1989). The lacI gene encodes a repressor of beta-galactosidase synthesis in *E. coli*. In a cell with functional lac repressor, the synthesis of beta-galactosidase is suppressed and colonies grown on X-gal plates are white. If the lac repressor gene is inactive, beta-galactosidase is produced and the colonies are a bright blue color. Because the function of the lac repressor can be measured with simple, in vivo assays it has been the subject of extensive genetic analysis (Markiewicz et al., J. Mol. Biol. 240:421-33, 1994; Suckow et al., J. Mol. Biol. 261:421-33, 1996). Based on this work, four G residues in a 205 base pair fragment which can not be changed without inactivating the protein were chosen. The sequence at these residues can thus be determined by assaying for Lac repressor function.

A 205 base pair segment of the lacI gene with the sequence:

```
1    AATTCATAAA GGAGATATCA TATGAAACCG GTAACGTTAT ACGACGTCGC TGAATACGCC
61   GGCGTTTCTT ACCAGACCGT TTCTAGAGTG GTTAACCAGG CTTCACATGT TAGCGCTAAA
121  ACCCGGGAAA AAGTTGAAGC TGCCATGGCT GAGCTCAACT ACATCCCGAA CCGTGTTGCG
181  CAGCAGCTGG CTGGTAAACA AAGCT
```

is synthesized using a set of overlapping double-stranded oligonucleotides.

The oligonucleotides used to make the gene are prepared using an Oligo 1000M DNA Synthesizer (Beckman Coulter, Inc, Fullerton, CA) using Beckman 30 nM DNA Synthesis Columns. All standard phosphoramidites and ancillary synthesis reagents are obtained from Glen Research, Inc. (Sterling, VA). Chemical phosphorylation of the oligonucleotides is done with the Chemical Phosphorylation II (Glen Research). Concentrated ammonia is obtained from Fisher Scientific (Springfield, NJ). 40% N-methylamine is obtained from Fluka Chemical Corporation (Milwaukee, WI). After cleavage from the solid support, the oligonucleotides are Trityl On purified using Poly-Pak Cartridges according to the instruction manual provided by Glen Research. Reagents for Trityl On purification are HPLC-grade acetonitrile and water obtained from Burdick & Jackson (Muskegon, MI). Triethylammonium acetate (TEAA), pH 7.0, and 3%

Trifluoroacetic acid in water are obtained from Glen Research. After purification, the synthesized oligonucleotides are evaporated to dryness in a SpeedVac (Savant, Farmingdale, NY) and resuspended in HPLC grade water. Concentrations of the oligonucleotides are determined by reading the 260 nm absorbance on a Pharmacia LKB Ultrospec III (Amersham Pharmacia, Upsala, Sweden).

The oligonucleotides are used to form duplex fragments by drying 500 pmoles each of the complementary oligonucleotides in a speedvac and resuspending in 10 microliters TE. A 5 microliter sample of the solution (250 pmoles) is mixed with 10 microliters of 2XSSPE (prepared according to Manniatis), heated to 95°C and cooled to room temperature.

Duplexes are successively ligated together to make longer fragments until the full length product is made. Each ligation consists of 500 picomoles of a pair of double-stranded oligonucleotide, 3 microliters of 10X ligation buffer (Fermentas Inc., Hanover, MD), 10 units of T4 DNA ligase (product # EL0016, Fermentas) and water to make a total volume of 30 microliters. All duplexes are ligated together under the same conditions. Each ligation mix is incubated at 37°C for 60 minutes, heated to 65°C for 10 minutes and the fragment isolated by HPLC.

High performance liquid chromatography (HPLC) is performed on a ProStar Helix HPLC system from Varian Inc. (Walnut Creek, CA) consisting of two high-precision high-pressure pumps (ProStar 215 Solvent Delivery Modules), a column oven (ProStar 510 Air Oven), a UV detector (ProStar 320 UV/Vis Detector) and a fraction collector (Dynamax FC-1 Fraction Collector), all controlled by Star Chromatography Workstation Software (Version 5.31). The column used is a Zorbax Eclipse dsDNA Analysis Column (4.6 mm ID x 75 mm, 3.5 micron) equipped with an in line Guard Column (4.6 mm ID x 12.5 mm, 3.5 micron) both from Agilent Technologies, Inc. (Palo Alto, CA). The following pre-made buffers are obtained from Varian Inc. (Walnut Creek, CA): Helix BufferPak "A" (100 mM Triethylammonium acetate, pH 7.0, 0.1 mM EDTA) and Helix BufferPak "B" (100 mM Triethylammonium acetate, pH 7.0, 0.1 mM EDTA with 25% by volume acetonitrile). The thermal and gradient conditions for isolating chemically-pure enriched sequence are calculated using the DHPLC Melt Program (http://insertion.stanford.edu/melt.html) available from Stanford University (Palo Alto, CA). Elution profiles are monitored using UV detector with absorbance at 260 nm.

The ligated fragments are dried down from the HPLC buffer and resuspended in TE. These fragments are used in a second set of ligation reactions. Several rounds of ligation followed by purification and fragment isolation are used to build the 205 base pair fragment of the lacI gene.

EXAMPLE 2

FUNCTIONAL TESTING OF THE 205 BASE PAIR FRAGMENT OF THE LACI GENE

The synthetic fragment produced in Example 1 is cloned into the lacI gene to test its function. Three micrograms of plasmid vector pWB1000 (Lehming et al., Proc. Natl. Acad. Sci. USA, 85:7947-7951, 1988) is digested with restriction enzymes EcoR1 and HindIII and the vector fragment gel purified using a Strata Prep DNA extraction kit (Stratagene product #400766) according to the manufacturers instructions, and resuspended in 100 microliters of TE. One microgram of the lacI fragment is treated with T4 polynucleotide kinase, extracted once with phenol and once with chloroform, ethanol precipitated and resuspended in 20 microliters of TE. Five microliters of the cut vector and one microliter of the synthetic lacI fragment are ligated in a total volume of 100 microliters using Fermentas T4 DNA ligase according to the manufacturers instructions. The ligation mix is extracted once with Strataclean, concentrated and washed twice with $1/10^{th}$ concentration TE and brought to a volume of 10 microliters in $1/10^{th}$ concentration TE. One microliter of this mix is transferred into *E. coli* strain DC 41-2 carrying plasmid pWB310 (Lehming et al., EMBO 6:3145-3153, 1987) by electroporation using a BTX ECM399 electroporator (Genetronics, Inc., San Diego, CA) according to the manufacturers instructions. Colonies are grown overnight on LB plates in the presence of 10 mg/liter tetracycline, 200 mg/liter ampicillin, 60 mg/liter X-gal and 300 mg/liter IPTG. Colonies carrying a plasmid with a functional lacI gene are white; those without a functional lacI gene are blue.

EXAMPLE 3

PREPARATION OF 205 BP DNA FRAGMENTS CONTAINING DIAMINOPURINE AT BASES 86, 88, 133, OR 178

One common side reaction of oligonucleotide synthesis is the formation of diaminopurine from a dG residue in the DNA chain. Modified oligonucleotides containing 2,6-diaminopurine are obtained from Trilink Biotechnologies (San Diego, CA) and incorporated into the 205 bp lacI gene fragment. Four samples are prepared as described in Example 1, with one diaminopurine residue (labeled **D** below) substituted for a dG residue in each sample.

Oligonucleotide	Fragment Name	Base Replaced
5′ ACCGTTTCTA**D**AGTGGTTAACCAGG 3′	D-T86	86
5′ ACCGTTTCTAGA**D**TGGTTAACCAGG 3′	D-T88	88
5′ GGAAAAA**D**TTGAAGCTGCCATGGCT 3′	D-T133	133
5′ TT**D**CGCAGCAGCTGGCTGGTAAACAA 3′	D-T178	178

EXAMPLE 4

PREPARATION OF 205 BP DNA FRAGMENTS CONTAINING A dU AT POSITIONS 86 OR 133

A second common side reaction of oligonucleotide synthesis is deamination of the N4-amine of deoxycytidine to form a uracil (dU) in the DNA chain. Modified oligonucleotides containing uracil (dU) are obtained from Midland Certified Reagent Company (Midland, TX) and incorporated into the 205 bp lacI gene fragment. Two samples are prepared as described in Example 1, with one uracil residue (labeled **dU** below) substituted for a dC residue in each sample.

Oligonucleotide	Fragment Name	Base Replaced
5′ TGAAGCCTGGTTAACCACT**dU**TAGAA 3′	U-B86	86
5′ AGCTCAGCCATGGCAGCTTCAA**dU**TT 3′	U-B133	133

EXAMPLE 5

PREPARATION OF 205 BP DNA FRAGMENTS CONTAINING AN ABASIC SITE AT POSITIONS 134 OR 182

A third common side reaction of oligonucleotide synthesis is the formation of abasic sites by depurination of protected adenosine residues during chain elongation. Modified oligonucleotides containing uracil are obtained from Midland Certified Reagent Company (Midland, TX) and incorporated into the 205 bp lacI gene fragment. Two samples are prepared as described in Example 1, with one uracil residue (labeled **dU** below) substituted for a dA residue in each sample.

Oligonucleotide	Fragment Name	Base Replaced
5' AGCTCAGCCATGGCAGCTTCA**dU**CTT 3'	A-B134	134
5' TTGCGC**dU**GCAGCTGGCTGGTAAACAA 3'	A-T182	182

After synthesis and HPLC purification of the 205 base pair fragments, the DNA is treated with Uracil-N-Glycosylase (Epicentre Technologies Corp., Madison, WI) according to the manufacturers instructions to remove the uracil base, leaving an apurinic site in place of the corresponding A residue in the native 205 base pair fragment.

EXAMPLE 6

CALCULATION OF THERMAL AND GRADIENT HPLC CONDITIONS FOR LACI SEQUENCE

The thermal and gradient conditions for isolating chemically-pure enriched sequence are calculated using the DHPLC Melt Program (http://insertion.stanford.edu/melt.html) available from Stanford University (Palo Alto, CA) and available for license from the Stanford University Office of Technology Licensing referring to the docket number S95-024. The 4 base single-stranded region on either end of the 205 base pair fragment is removed to give the following 197 base pair sequence.

lac I Region

CATAAAGGAGATATCATATGAAACCGGTAACGTTATACGACGTCGCTGAA
TACGCCGGCGTTTCTTACCAGACCGTTTCTAGAGTGGTTAACCAGGCTTC
ACATGTTAGCGCTAAAACCCGGGAAAAAGTTGAAGCTGCCATGGCTGAGC
TCAACTACATCCCGAACCGTGTTGCGCAGCAGCTGGCTGGTAAACAA

The gradients are specified below as percent buffer B at times 1, 2 and 3 (B1, B2, B3). The gradient is run from B1 to B2 in 0.5 minutes, then B2 to B3 in 3.0 minutes.

Conditions	Temperature (C)	B1	B2	B3
1	53	50	59.6	65
2	55	50	56.8	62.2
3	57	50	54.1	59.5
4	59	50	51.4	56.8
5	61	45	50	55.4

Buffer A and buffer B are as described in Example 1.

EXAMPLE 7

DETERMINATION OF THE TEMPERATURE-DEPENDENT CHROMATOGRAPHIC PROFILES OF THE NATIVE AND EIGHT MODIFIED LACI FRAGMENTS

The chromatographic behavior of the native lacI DNA and the eight modified lacI DNA are measured in response to a range of gradient and temperature conditions. The lacI DNA is below:

Name	Type and Location of Modification
Pure	No chemical modification
D-T86	2,6-diaminopurine @ position 86
D-T88	2,6-diaminopurine @ position 88
D-T133	2,6-diaminopurine @ position 133
D-T178	2,6-diaminopurine @ position 178
U-B86	2'-deoxyuridine @ position 79
U-B133	2'-deoxyuridine @ position 133
A-B134	abasic @ position 134
A-T182	abasic @ position 182

25 pmoles of each sample is suspended in 5 µl of HPLC-grade water and directly chromatographed on a Zorbax Eclipse ds DNA Analysis Column (4.6 mm ID x 75 mm, 3.5 micron) with an in line Pre-Column (4.6 mm ID x 12.5 mm, 3.5 micron) with Buffer A consisting of 100 mM Triethylammonium acetate, pH 7.0, 0.1 mM EDTA and Buffer B consisting of 100 mM Triethylammonium acetate, pH 7.0, 0.1 mM EDTA with 25% by volume acetonitrile. The details of each gradient and temperature condition are as described in Example 6.

Each fragment denatures at a temperature that is a function of the strength of the duplex structure. The fully base paired native lacI sequence forms the most stable duplex and denatures under the most stringent conditions. Fragments with base modifications form less stable duplexes, denature at a lower temperature and thus show earlier elution at a given temperature and gradient profile.

EXAMPLE 8

ENZYMATIC AMPLIFICATION OF A CHEMICALLY-SYNTHESIZED 205 BASE PAIR LACI GENE CARRYING MODIFIED BASES

The fragments produced in Example 3, Example 4 and Example 5 (fragments D-T86, D-T88, D-T133, D-T178, U-B86, U-B133, A-B134, A-T182) are amplified using PCR to convert the base pair mismatches in the synthetic fragments into base paired errors in the enzymatically produced fragments. The PCR is designed to amplify the complete fragment and add sequence using tails on the primers to add cloning sites for EcoR1 and HindIII restriction enzymes.

Tailed PCR Primer Sequences

Forward primer: 5′-AGGCTGAAACTGGACAATTCATAAAGGAGATATCATATGAAACCG-3′

Reverse primer: 5′-CTTCGGAAGATCCTTAGCTTTGTTTACCAGCCAGCTG-3′

The PCR conditions for amplification of the fragments produced in Example 3, Example 4 and Example 5 are described in the table below. All the components are combined and vortexed to ensure good mixing, and centrifuged. Aliquots are then distributed into PCR tubes as shown in the following table:

COMPONENT	VOLUME
Pfu 10X Buffer (Cat. No.600153-82, Stratagene, Inc., La Jolla, CA)	5 μL
10 mM dNTP Mix	1 μL
Forward primer (10 pmol/μL)	1 μL
Reverse primer (10 pmol/μL)	1 μL
H_2O	36 μL
Synthetic DNA Fragment	5 μL
PFUTurbo™ (Cat. No. 600250, Stratagene, Inc., La Jolla, CA)	1 μL

The PCR tubes are placed into a thermocycler (MJ Instruments) and the temperature cycling program was initiated. The cycling program parameters are shown in the table below:

STEP	TEMP	TIME
1	94C	2 minutes
2	94C	1 minute
3	55C	1 minute
4	72C	1 minute
5	Go to Step 2, 30X	
6	72C	10 minutes
7	4C	Forever

The PCR products are purified using the QIAquick PCR Purification Kit (Qiagen Inc., Valencia, CA) according to the manufacturers instructions and resuspended in 10 microliters of TE.

EXAMPLE 9

FUNCTIONAL TEST OF 205 BASE PAIR ENZYMATICALLY AMPLIFIED FRAGMENTS OF THE LACI GENE CARRYING MODIFIED BASES

The enzymatic fragments produced in Example 8 are cloned into the lacI gene to test their biological function. Ten micrograms of plasmid vector pWB1000 (Lehming et al., PNAS 85:7947-7951, 1988) and each of the PCR reactions from Example 8 is digested with restriction enzymes EcoR1 and HindIII . Each of the cut amplification products and the vector fragment are gel purified using a Strata Prep DNA extraction kit (Stratagene product #400766) according to the manufacturers instructions, and resuspended in 100 microliters of TE. Each of the cut PCR reactions and one microgram of each lacI fragment is treated with T4 polynucleotide kinase, extracted once with phenol and once with chloroform, ethanol precipitated and resuspended in 20 microliters of TE. Five microliters of the cut vector and the entire sample of the amplified DNA are ligated in a total volume of 100 microliters using New England Biolabs T4 DNA ligase according to the manufacturers instructions. The ligation mix is extracted once with Strataclean, concentrated and washed twice with $1/10^{th}$ concentration TE and brought to a volume of 10 microliters in $1/10^{th}$ concentration TE. One microliter of this mix is transferred into *E. coli* strain DC 41-2 carrying plasmid pWB310 (Lehming et al., EMBO 6:3145-3153, 1987) by electroporation using a BTX ECM399 electroporator according to the manufacturers instructions. Colonies are grown overnight on LB plates in the presence of 10 mg/liter tetracycline, 200 mg/liter ampicillin, 60 mg/liter X-gal and 300 mg/liter IPTG. Colonies carrying a plasmid with a functional lacI gene are white; those without a functional lacI

gene are blue. Each modified fragment is characterized by the frequency of blue colonies relative to the frequency of blue colonies derived from clones of the native synthetic lacI fragment as described in Example 2.

EXAMPLE 10

ENRICHMENT OF NATIVE LACI FRAGMENTS FROM MIXTURES OF NATIVE AND ENZYMATICALLY AMPLIFIED FRAGMENTS OF THE LACI GENE CARRYING MODIFIED BASES BY PREPARATIVE HPLC

The ability of the chromatographic technique to enrich a population of enzymatically amplified base paired DNA composed of "correct" DNA in the presence of "incorrect" DNA is shown by spiking native lacI DNA with each of the eight amplified lacI DNA from Example 8, denaturing and reannealing the mix, and enriching for the correct DNA using HPLC. For each of the eight amplified DNA fragments from Example 8 an equimolar mixture is prepared of amplified native and amplified fragments by mixing 20 pmoles of the amplified DNA with 20 pmoles of the amplified native fragment. A fraction of each mixture is retained for functional testing as described below. The remainder of each of these samples is chromatographed using thermal and gradient conditions (identified in Example 7) which alter the mobility of the modified fragments relative to the native fragment. For each sample, the peaks are collected with a fraction collector as described in Example 1 at the elution time determined in Example 7. Two fractions are collected, one with a mobility characteristic of the modified DNA fragments and one with a slower mobility characteristic of the native DNA fragment. These fractions are dried down and cloned as described in Example 9. In parallel, a portion of each of the eight unfractionated mixtures is cloned and tested in the same way. The "native fraction" fragments show a lower number of sequence errors than the original mixtures or the early-eluting fractions, as indicated by the frequency of blue colonies.

EXAMPLE 11

ENRICHMENT OF NATIVE LACI FRAGMENTS FROM MIXTURES OF NATIVE AND ENZYMATICALLY AMPLIFIED FRAGMENTS OF THE LACI GENE CARRYING MODIFIED BASES BY REMOVAL OF MISMATCHES WITH A MISMATCH BINDING PROTEIN IMMOBILIZED TO MAGNETIC BEADS

The ability of the mismatch binding protein to enrich a population of enzymatically amplified base paired DNA composed of "correct" DNA in the presence of "incorrect" DNA is shown by spiking native lacI DNA with each of the eight amplified lacI DNA from Example 8, denaturing and reannealing the mix, and enriching for the correct

DNA using mismatch binding protein immobilized to magnetic beads. For each of the eight amplified DNA fragments from Example 8 an equimolar mixture is prepared of amplified native and amplified fragments by mixing 20 pmoles of the amplified DNA with 20 pmoles of the amplified native fragment. A fraction of each mixture is retained for functional testing as described below.

The remainder of each of these samples are exposed to MutS immobilized on magnetic beads (GeneCheck Inc., Fort Collins, CO). The magnetic particles are collected on the walls of the tube with a magnet and the supernatant is collected. These supernatants are dried down and cloned as described in Example 9. In parallel, a portion of each of the eight unpurified mixtures are cloned and tested in the same way. The "native fraction" fragments show a lower number of sequence errors than the original mixtures or the early-eluting fractions, as indicated by the frequency of blue colonies.

EXAMPLE 12

ENRICHMENT OF NATIVE LACI FRAGMENTS FROM MIXTURES OF NATIVE AND ENZYMATICALLY AMPLIFIED FRAGMENTS OF THE LACI GENE CARRYING MODIFIED BASES BY REMOVAL OF MISMATCHES WITH A MISMATCH BINDING PROTEIN PASSED THROUGH A NITROCELLULOSE FILTER

The ability of the mismatch binding protein to enrich a population of enzymatically amplified base paired DNA composed of "correct" DNA in the presence of "incorrect" DNA is shown by spiking native lacI DNA with each of the eight amplified lacI DNA from Example 8, denaturing and reannealing the mix, and enriching for the correct DNA using mismatch binding protein passed through a nitrocellulose filter. For each of the eight amplified DNA fragments from Example 8 an equimolar mixture is prepared of amplified native and amplified fragments by mixing 20 pmoles of the amplified DNA with 20 pmoles of the amplified native fragment. A fraction of each mixture is retained for functional testing as described below.

The remainder of each of these samples are exposed to MutS (Amersham Pharmacia Biotech, Upsala, Sweden) immobiled on to a nitrocellulose filter. A nitrocellulose sheet (0.45 micron, Schleicher and Schull, BA85) was pre-wet by floating in reaction buffer (20 mM Tris HCl, pH 7.6; 5 mM $MgCl_2$, 0.1 mM DTT, 0.01 mM EDTA). MutS (500 ng/well) in reaction buffer was bound to the nitrocellulose in a 48 well slot blotting apparatus (Hoefer Scientific Instruments) over 3 sheets of blotting paper (Schleicher and Schull GB002). After 20 min. at room temperature, nitrocellulose was blocked with 200 µl/well of 3% horse radish peroxidase (HRP)-free bovine serum albumin (BSA). After 1 hour, excess blocking solution was pulled through with vacuum and DNA

(1 ng and 10 ng) was added in 20 µl reaction buffer containing 3% BSA. After 30 min. at room temperature, wells were washed 1 time with 100 µl reaction buffer. The wash fluids were decanted rather than pulled through with vacuum. These supernatants are dried down and cloned as described in Example 9. In parallel, a portion of each of the eight unpurified mixtures are cloned and tested in the same way. The "native fraction" fragments show a lower number of sequence errors than the original mixtures or the early-eluting fractions, as indicated by the frequency of blue colonies.

From the foregoing it will be appreciated that, although specific embodiments of the invention have been described herein for purposes of illustration, various modifications may be made without deviating from the spirit and scope of the invention.

CLAIMS

1. A method of depleting in a sample of double-stranded oligonucleotides a population of double-stranded oligonucleotides containing mismatched bases thereby enriching in said sample a population of double-stranded oligonucleotides containing correctly matched bases, comprising the steps of:

(a) contacting said sample containing double-stranded oligonucleotides with a mismatch recognition protein under conditions to permit the protein to interact with a double-stranded oligonucleotide containing at least one mismatched base; and

(b) collecting double-stranded oligonucleotides that have not interacted with said mismatch recognition protein, thereby depleting the population of double-stranded oligonucleotides containing mismatched bases.

2. The method of claim 1, wherein prior to the step of collecting, having an additional step comprising separating said double-stranded oligonucleotide containing at least one mismatched base that has interacted with said mismatch recognition protein, from double-stranded oligonucleotides that have not interacted with said mismatch recognition protein.

3. The method of claim 1 wherein the double-stranded oligonucleotides of said sample are chemically synthesized.

4. The method of claim 1 wherein the double-stranded oligonucleotides of said sample are enzymatically synthesized.

5. The method of claim 4, wherein prior to the step of contacting, having additional steps comprising denaturing and reannealing said sample of double-stranded oligonucleotides under conditions to permit conversion of the double-stranded oligonucleotides first to single-stranded oligonucleotides and then to double-stranded oligonucleotides.

6. The method of claim 1 wherein said mismatch recognition protein is immobilized on a solid support.

7. The method of any one of claims 1, 2, 3, 4, 5 or 6 wherein said double-stranded oligonucleotides are double-stranded DNA.

8. The method of claim 7 wherein the DNA is a gene or a portion of a gene.